

Original citation:

Cheng, Long, Zhang, Huaizhi, Jose , Joemon M., Yu, Haitao, Moshfeghi, Yashar and Triantafillou, P.. (2017) Topic detection and tracking on heterogeneous information. Journal of Intelligent Information Systems.

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/93997>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 4.0 International license (CC BY 4.0) and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by/4.0/>

A note on versions:

The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk



Topic detection and tracking on heterogeneous information

Long Chen¹  · Huaizhi Zhang¹ · Joemon M Jose¹ ·
Haitao Yu¹ · Yashar Moshfeghi² · Peter Triantafillou¹

Received: 16 August 2016 / Revised: 5 September 2017 / Accepted: 6 September 2017
© The Author(s) 2017. This article is an open access publication

Abstract Given the proliferation of social media and the abundance of news feeds, a substantial amount of real-time content is distributed through disparate sources, which makes it increasingly difficult to glean and distill useful information. Although combining heterogeneous sources for topic detection has gained attention from several research communities, most of them fail to consider the interaction among different sources and their intertwined temporal dynamics. To address this concern, we studied the dynamics of topics from heterogeneous sources by exploiting both their individual properties (including temporal features) and their inter-relationships. We first implemented a heterogeneous topic model that enables topic–topic correspondence between the sources by iteratively updating its topic–word distribution. To capture temporal dynamics, the topics are then correlated with a time-dependent function that can characterise its social response and popularity over time. We extensively evaluate the proposed approach and compare to the state-of-the-art techniques on heterogeneous collection. Experimental results demonstrate that our approach can significantly outperform the existing ones.

Keywords Topic detection · Heterogeneous sources · Temporal dynamics · Social response · Topic importance

✉ Long Chen
long.chen@glasgow.ac.uk

Yashar Moshfeghi
yashar.moshfeghi@strath.ac.uk

¹ School of Computing Science, University of Glasgow, Glasgow, UK

² School of Computing Science, University of Strathclyde, Glasgow, UK

1 Introduction

Social media, such as Twitter and Facebook, has been used widely for communicating breaking news, eye witness account, and even organising flash mobs. Users of these websites have become accustomed to receiving timely updates on important events. For example, Twitter was heavily used in numerous international events, such as the Ukrainian crisis (2014) and the Malaysia Airlines Flight 370 crisis (2015). From a user consumer's perspective, it makes sense to combine social media with traditional news outlets, e.g., BBC news, for timely and effective news consumption. However, the latter have different temporal dynamics than the former, which entails a deep understanding of the interaction between the new and old sources of news.

Combining heterogeneous sources of news has been investigated by several research communities (cf. Section 2). However, existing works mostly merge documents from all sources into a single collection, and then apply topic modelling techniques to it for detecting common topics. This may cause a biased result in favour of the source with a high frequencies of publication. Furthermore, the heterogeneity that characterises each source may not be maintained. For example, the Twitter data stream is distinctively biased towards current topics and temporal activity of users, this means for effective topic modelling, computational treatment of user's social behaviours (such as author information and the number of retweets) is also needed. Alternatively, running the existing topic models on each source can preserve the characteristics of each source, but make it difficult to capture a common topic distribution and the interaction among different sources. To solve the aforementioned problems, we present a heterogeneous topic model that can combine multiple disparate sources in a complementary manner by assuming a variable of common-topic distribution for both Twitter and news collections.

Furthermore, people would not only like to know the type of topic that can be found from these disparate data sources but also desire to understand their temporal dynamics, as well as the topic importance. However, the dynamics of most topic streams are intertwined with each other across sources such that their impact is not easily recognisable. To determine the impact of a topic, it is critical to consider the evolution of the aggregated social response from social media (i.e., Twitter), in addition to the temporal dynamics of news media. However, news media can have the news cycle (Tsytarau et al. 2014) all by themselves while keeping a growth shape, such that the burst shape of publication does not always consistent with the beginning of the topic. Therefore, we used a deconvolution approach (a well-known technique in audio and signal processing (Kirkeby et al. 1998; Mallat 1999)) that can address these concerns by using a special compound function that considers both topic importance and its social response in social media.

In this study, we addressed the problem of *Topic Detection and Tracking* (TDT) from heterogeneous sources. Our study is based on recent advances in both topic modelling and information cascading in social media. In particular, we designed a heterogeneous topic model that allows information from disparate sources to communicate with each other while maintaining the properties of each source in a unified framework. For temporal modelling, we proposed a compound function through convolution, which optimally balances the topic importance and its social response. By combining these two models, we effectively modelled temporal dynamics from disparate sources in a principled manner.

The rest of the paper is organised as follows. Section 2 describes background and related work. In Sections 3 and 4, we discuss our model in details. Section 5 provides experimental results on two real-world datasets. We present our conclusion and future work in Section 6.

2 Related work

There are approaches that tackle sub-tasks of our problem in various domains, however, they cannot be combined to solve the problem that our model solves. To the best of our knowledge, there is no existing work that is capable of automatically detecting and tracking topics from heterogeneous sources while simultaneously preserving the properties of each source. The task of this paper can be loosely organised into two independent sub-tasks: (1) topic detection from heterogeneous sources (2) characterising their temporal dynamics. In this section, we will review these two lines of related work.

2.1 Topic detection from heterogeneous sources

One of the track tasks included in the Topic Detection and Tracking (TDT) (Fiscus and Doddington 2002) is topic detection, where systems cluster streamed stories into bins depending on the topics being discussed. The techniques of topic modelling, such as PLSA (Hofmann 2001) and LDA (Blei et al. 2003), have shown to be effective for topic detection. In the seminal work of online-LDA model (Alsumait et al. 2008), the authors update the LDA incrementally with information inferred from the new stream of data. Lau et al. (2012) subsequently demonstrated the effectiveness of Online-LDA for Twitter data streams.

However, there has not been extensive research on the development of topic detection from heterogeneous sources. Zhai et al. (2004) proposed a cross-collection mixture model to detect common topics and local ones respectively. The state-of-the-art approach (Hong et al. 2011) utilises the collection model for mining from multiple sources, together with a meme-tracking model that iteratively updates the hyper-parameter that controls the document-topic distribution to capture the temporal dynamics of the topic. In the Collection Model, a word belongs to either the local topic or the common topic, the probability of which is drawn from a Bernoulli distribution. Common topics are obtained by merging all documents from all sources into one single collection, and then apply LDA on it. Local topics are computed by using LDA on each source individually. But it assumes that there is no correspondence between local topics across different sources, nor is there exchanging information between the local topics and the common ones. While in real-world scenarios, information from multiple sources constantly interacted with each other as the topic evolves. To address this problem, Ghosh and Asur (2013) proposed a Source LDA model to detect topics from multiple sources with the aim to incorporate source interactions, which is somewhat similar to our idea. However, there are stark differences between their work and ours: First, their model assumes that there is no order for the documents in the collection, hence the temporal dynamics of each source is completely ignored. Secondly, since the same LDA model is applied over different sources, they didn't exploit the properties that characterise each data source, whereas we use Author Topic model and LDA for Tweets and News sources respectively (cf. Section 3).

2.2 Characterizing temporal dynamics and social response

In the seminal work of dynamic topic modelling (Blei and Lafferty 2006), each topic defines a multinomial distribution over a set of terms. Therefore, for each word of each document, a topic is drawn from the mixture and a term is subsequently drawn from the multinomial distribution corresponding to that topic. This has led to the recent development of incorporating temporal dynamics into topic models (e.g., Wang et al. 2012; Hong et al. 2011; Masada et al. 2009; Wang and McCallum 2006; Dubey et al. 2013). These models enable

us to gain insight into datasets with temporal changes in a convenient way and open future directions for utilizing these models in a more general fashion. However, their analysis was conducted on an academic dataset. To investigate the effectiveness of dynamic topic model, Leskovec et al. (2009) proposed a framework to capture the textual variants over different phrases. It is based on two assumptions for the interaction of Memes sources: imitation and recency. The imitation hypothesis assumes that news sources are more likely to publish on events that have already seen large volume of publications. The recency hypothesis marks the tendency to publish more on recent events. While effective, their work only focused on the temporal dynamics of one source, namely the news outlets, whereas in this paper two different types of sources are considered simultaneously.

In addition, while there are previous work investigate the temporal dynamics in a heterogeneous context (Hong et al. 2011), a deeper understanding of topic dynamics entails extraction of burst shapes and modelling of social response (Tsytarau et al. 2014). This requirement is important as publication volume often contains background information, which may mask individual patterns of topics. Another important factor is topic importance, which is first proposed in Cha et al. (2010). The authors studied the influence within

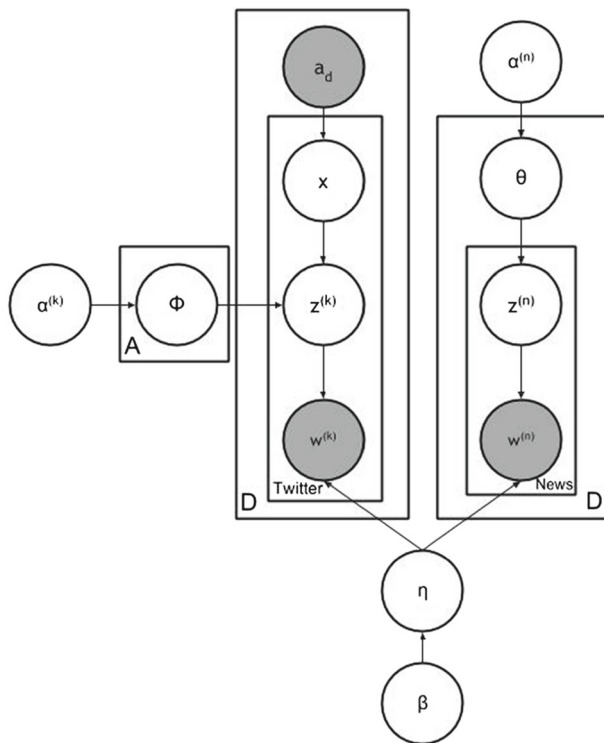


Fig. 1 The framework of heterogeneous topic model (cf. Table 1)

Twitter, and performed a comparison of three different measures of influence- indegree, retweets, and user mentions. They discovered that retweets and user mentions are better measure of topic influence than indegree. Unlike the previous studies, in this paper, we aim to incorporate both topic importance and social response into a unified topic modelling framework.

3 Heterogenous topic model

In this section, we will explain the heterogeneous topic model, which can detect topics in heterogeneous sources while preserving the properties of each source. Throughout the paper, we have used the general term “document” or “feed” to cover the basic text collections. For example, in the context of news media, a feed is a news article; whereas for Twitter, a feed represents a tweet message.

Table 1 Notations used in HTM

Symbol	Description
N	Number of words in the collection(Twitter and News)
N_d	Number of common words appeared in both collections
W	Vocabulary size
T	Number of topics
A	Number of authors in Twitters
C^{WT}	Number of words assigned to the topic of a word
C^{TA}	Number of words assigned to the topic of an author
$n_d^{(t)}$	Frequency that topic t assigned to a word in document d
$n_t^{(w)}$	Frequency that word w assigned to topic t
$w_{di}^{(k)}$	Words in Twitter document d
$w_{di}^{(n)}$	Words in News document d
z	Topic assignment
z_{di}	Topic assignment for word w_{di}
x	Author assginments
x_{di}	Author assignment for word w_{di}
\mathcal{A}	Authors of the corpus in Twitters
$\alpha^{(k)}$	Dirichlet prior for Twitter
$\alpha^{(n)}$	Dirichlet prior of News document
α_t	Dirichlet prior for topic t in News document
β	Dirichlet prior
β_w	Dirichlet prior of word w
η	Probabilities of words given on topics

3.1 Model description

To correctly model the topics and the distinct characteristics of both Twitter and Newswire, we propose Heterogeneous Topic Model, abbreviated as HTM. Our model can identify topics across disparate sources while preserving the properties of each source. For each source, any local topic i of any source j would correspond to topic i of another source k , where topic i is conformed to the properties of the local source. With the notation given in Table 1, Fig. 1 illustrates the graphical model for HTM, which blends two topic modelling techniques, namely, Author Topic Model (ATM) (Rosen-Zvi et al. 2004) and Latent Dirichlet Allocation (LDA).

The various probability distributions we can learn from the HTM model characterise the different factors of each source that can affect the topics. For the generation of content, each word $w^{(n)}$ in news articles is only associated with a topic z , and each word $w^{(k)}$ in tweets is related to two latent variables, namely, an author a and a topic z . The communication between the topics of different sources is governed by a parameter, η , which represents the common topic distribution. β is the prior distribution of η . The observed variables include author names of tweets, words in the Twitters dataset, and words in the News dataset, the rest are all unobserved variables. Notice that the D in the left side of the figure represent the Twitter documents and the D in the right side of the figure is the News documents.

Conditioned on the set of authors from Twitters and its distribution over topics, the generative process of the HTM are summarized in Algorithm 1, where $variable^{(k)}$ represents the variable of Twitters and $variable^{(n)}$ denotes the variable of news. The local words represents the ones that only appeared in a single source, whereas the common words are the ones occurred in all the sources. Under the generative process, each common topic z in News and Twitters is drawn independently when conditioned on Θ and Φ , respectively.

Holding the conditional independence property, we have the following basic equation for Gibbs sampler:

$$\begin{aligned}
 p(z_{di,dj} = t | w_{di}^{(k)} = w_i, z_{-di}, x_{-di}, w_{-di}^{(k)}, \\
 w_{dj}^{(n)} = w_j, z_{-dj}, w_{-dj}^{(n)}, \mathcal{A}, \alpha^{(k)}, \alpha^{(n)}, \beta) \\
 \propto p(z_{di,dj} = t, w_{di}^{(k)} = w_i, w_{dj}^{(n)} = w_j | z_{-di}, x_{-di}, \\
 w_{-di}^{(k)}, z_{-dj}, w_{-dj}^{(n)}, \mathcal{A}, \alpha^{(k)}, \alpha^{(n)}, \beta) \\
 = \frac{p(z, w^{(k)}, w^{(n)} | \mathcal{A}, \alpha^{(k)}, \alpha^{(n)}, \beta)}{p(z_{-di}, z_{-dj}, w_{-di}^{(k)}, w_{-dj}^{(n)} | \mathcal{A}, \alpha^{(k)}, \alpha^{(n)}, \beta)} \\
 = \frac{p(z, w^{(k)} | \mathcal{A}, \alpha^{(k)}, \beta)}{p(z_{-di}, w_{-di}^{(k)} | \mathcal{A}, \alpha^{(k)}, \beta)} \cdot \frac{p(z, w^{(n)} | \alpha^{(n)}, \beta)}{p(z_{-dj}, w_{-dj}^{(n)} | \alpha^{(n)}, \beta)}
 \end{aligned} \tag{1}$$

where z_{-di} , w_{-di} stand for the vector of topic assignments and word observations except for the i^{th} word of news document d . z_{-dj} , w_{-dj} stand for the vector of topic assignments and word observations except for the j^{th} word of tweet d .

Algorithm 1 The generative process of the heterogeneous topic model

```

1 for each topic  $t = 1, \dots, T$ , author of the tweet  $a = 1, \dots, A$  do
2   choose  $\eta_t \sim \text{Dirichlet}(\beta)$ 
3   choose  $\phi_a \sim \text{Dirichlet}(\alpha^{(k)})$ 
4 end
5 given the vector of authors  $a_d$  of Twitter  $d_i^{(k)}$ :
6 for each common word  $i = 1, \dots, N_d$  in News and Twitter do
7   choose a topic  $z_{di}^{(n)} \sim \text{Multinomial}(\theta)$ 
8   choose a word  $w_{di}^{(n)} \sim \text{Multinomial}(z_{di}^{(n)}, \eta_{z_{di}^{(n)}})$ 
9   choose an author  $x_{di} \sim \text{Uniform}(a_d)$ 
10  choose a topic  $z_{di}^{(k)} \sim \text{Multinomial}(\phi_{x_{di}}, \theta)$ 
11  choose a word  $w_{di}^{(k)} \sim \text{Multinomial}(\eta_{z_{di}^{(k)}})$ 
12 end
13 for each document in news collection  $d = 1, \dots, D$  do
14   for each local word  $i = 1, \dots, N_d$  do
15     choose a topic  $z_{di}^{(n)} \sim \text{Multinomial}(\theta)$ 
16     choose a word  $w_{di}^{(n)} \sim \text{Multinomial}(z_{di}^{(n)}, \eta_{z_{di}^{(n)}})$ 
17   end
18 end
19 for each document in twitter collection  $d = 1, \dots, D$  do
20   given the vector of authors  $a_d$ 
21   for each local word  $i = 1, \dots, N_d$  do
22     choose an author  $x_{di} \sim \text{Uniform}(a_d)$ 
23     choose a topic  $z_{di}^{(k)} \sim \text{Multinomial}(\phi_{x_{di}}, \theta)$ 
24     choose a word  $w_{di}^{(k)} \sim \text{Multinomial}(\eta_{z_{di}^{(k)}})$ 
25   end
26 end

```

After integrating the joint distribution of the variables (cf. [Appendix](#)), we get the following equation for Gibbs sampler:

$$\begin{aligned}
 p(z_{di,dj} = t | w_{di}^{(k)} = w_i, z_{-di}, x_{-di}, w_{-di}^{(k)}, \\
 w_{dj}^{(n)} = w_j, z_{-dj}, w_{-dj}^{(n)}, \mathcal{A}, \alpha^{(k)}, \alpha^{(n)}, \beta) \\
 = \frac{C_{wt,-di}^{WT} + \beta_w}{\sum_{w'} C_{w't,-di}^{WT} + W\beta} \cdot \frac{C_{ta,-di}^{TA} + \alpha^{(k)}}{\sum_{t'} C_{t'a,-di}^{TA} + T\alpha^{(k)}} \\
 \cdot \frac{n_{t,-i}^{(w)} + \beta_w}{\sum_{w=1}^V n_{t,-i}^{(w)} + \beta_w} \cdot \frac{n_{d,-t}^{(t)} + \alpha_t}{[\sum_{t=1}^T n_d^{(t)} + \alpha_t] - 1}
 \end{aligned} \quad (2)$$

Note that, the sampled words from one collection may not be observed in another collection. In such cases, the prior probability of topic over word is set as one.

3.2 Model fitting via Gibbs sampling

In order to estimate the hidden variables of HTM, we use collapsed Gibbs Sampling. However, the derivation of posterior distributions for Gibbs sampling in HTM is complicated by the fact that common distribution η is a joint distribution of two mixtures. As a result, we need to compute the joint distribution of $w^{(k)}$ and $w^{(n)}$ in the Gibbs sampling process. The posterior distributions for Gibbs sampling in HTM are

$$\beta_t^{(n)} | \mathbf{z}, \mathcal{D}^{\text{train}}, \beta \sim \text{Dirichlet} \left(C_t^{WT} + (C_t^{WT})^{(n)} + \beta \right) \quad (3)$$

$$\beta_t^{(k)} | \mathbf{z}, \mathcal{D}^{\text{train}}, \beta \sim \text{Dirichlet} \left(C_t^{WT} + (C_t^{WT})^{(k)} + \beta \right) \quad (4)$$

$$\beta_t | \mathbf{z}, \mathcal{D}^{\text{train}}, \beta \sim \text{Dirichlet} (C_t^{WT} + (C^{WT})^{(n)} + (C^{WT})^{(k)} + \beta) \quad (5)$$

$$\phi_a | \mathbf{x}, \mathbf{z}, \mathcal{D}^{\text{train}}, \alpha^{(k)} \sim \text{Dirichlet} (C_a^{TA} + \alpha^{(k)}) \quad (6)$$

$$\theta_d | \mathbf{w}, \mathbf{z}, \mathcal{D}^{\text{train}}, \alpha^{(n)} \sim \text{Dirichlet} (n_d + \alpha^{(n)}) \quad (7)$$

where $\beta_t^{(n)}$ represents the local topics that belong to News; $(C^{WT})^{(n)}$, $(C^{WT})^{(k)}$ indicates the sample of topic-term matrix in which each term is observed only in Twitter and only in News respectively, (C^{WT}) is the sample of topic-term matrix in which each word can be observed from the collections; $\beta_t^{(t)}$ is the local topics that belong to Twitter. Since the Dirichlet distribution is conjugate to the Multinomial distribution, the posterior mean of \mathcal{A} , Θ and Φ given $\mathbf{x}, \mathbf{z}, \mathbf{w}, \mathcal{D}^{\text{train}}, \alpha^{(n)}, \alpha^{(k)}$ and β can be obtained as follows:

$$E[\beta_{wt}^{(n)} | \mathbf{z}^s, \mathcal{D}^{\text{train}}, \beta] = \frac{(C_{wt}^{WT} + (C_{wt}^{WT})^{(n)})^s + \beta_w}{\sum_{w'} (C_{w't}^{WT} + (C_{w't}^{WT})^{(n)})^s + W\beta} \quad (8)$$

$$E[\beta_{wt}^{(k)} | \mathbf{z}^s, \mathcal{D}^{\text{train}}, \beta] = \frac{(C_{wt}^{WT} + (C_{wt}^{WT})^{(k)})^s + \beta_w}{\sum_{w'} (C_{w't}^{WT} + (C_{w't}^{WT})^{(k)})^s + W\beta} \quad (9)$$

$$E[\beta_{wt} | \mathbf{z}^s, \mathcal{D}^{\text{train}}, \beta] = \frac{(C_{wt}^{WT} + (C_{wt}^{WT})^{(n)} + (C_{wt}^{WT})^{(k)})^s + \beta_w}{\sum_{w'} (C_{w't}^{WT} + (C_{w't}^{WT})^{(n)} + (C_{w't}^{WT})^{(k)})^s + W\beta} \quad (10)$$

$$E[\phi_{ta} | \mathbf{z}^s, \mathbf{x}^s, \mathcal{D}^{\text{train}}, \alpha^{(k)}] = \frac{(C_{ta}^{TA})^s + \alpha^{(k)}}{\sum_{t'} (C_{t'a}^{TA})^s + T\alpha^{(k)}} \quad (11)$$

$$E[\theta_{dt} | \mathbf{w}^s, \mathbf{z}^s, \mathcal{D}^{\text{train}}, \alpha^{(n)}] = \frac{(n_d^{(t)})^s + \alpha_t}{\sum_{t=1}^T (n_d^{(t)})^s + \alpha_t} \quad (12)$$

where s refers to the sample from Gibbs sampler of the full collection. The posterior \mathcal{A} , Θ and Φ , correspond to the author distribution on topics, the topic distribution on words, and the document distribution on topics respectively.

4 Modelling temporal dynamics

In this section, we review a temporal model for topics, introduced in Tsytarau et al. (2014) and present an alternate derivation. We start with the description of the basic social

response function and our representation of topic importance. Then we introduce a deconvolution approach over the time series of news and Twitter data, in order to extract important properties of topics.

4.1 Modelling impacting topics

As one may notice, not every publications outbursts is derived from external stimuli. For example, there are two different types of dynamics in Twitter: daily activity and trending activity. The former is mostly driven by work schedules of time zones and the latter is caused by a more clear pattern of topic interest and is the subject of our study. We start by assuming the following setting, which assumes the observed topic dynamics (volume of publications) as a response of social media and topic importance. The result is decomposed into two functions: the topics importance function and the social media response function:

$$n(t) = \int_{-\infty}^{+\infty} srf(\tau)e(t - \tau, r)d\tau \quad (13)$$

where $srf(t)$ is the social response function; and $e(t, r)$, which reflects the importance of the topic during time t , is the joint function of actual topic sequence t and the number of retweets r . $e(t, r) = \ln(r)(a + b)t_0 - bt, t > t_0$; $e(t, r) = \ln(r)at, 0 < t < t_0$. a is buildup rate and b is decay rate. The intuition behind $e(t, r)$ is that certain topics should have a better chance of being selected, since they had a higher popularity and a larger volume of social response. For example, during the "heat wave action", the hottest ever day of UK in twelve years, news articles and Twitter messages are more inclined to talk about weather, rather than politics. Furthermore, we observe that hot topics usually correspond to a higher retweet number. The form of (13) has been demonstrated to be effective for capturing spikes of news articles and Tweets (Hong et al. 2011).

However, in order to restore the original topic sequence, it is important to know the exact shape of $srf(t)$. To model the shape of $srf(t)$, we propose to employ a family of normalized decaying functions (Asur et al. 2011) demonstrated in the following equations:

$$\text{linear} \quad srf(t) = \left(\frac{2}{\tau_0} - \frac{2t}{\tau_0}\right)h(t)h(\tau_0 - t) \quad (14)$$

$$\text{hyperbolic} \quad srf(t) = h(t)\frac{\alpha - 1}{\tau_0} \frac{t + \tau_0}{\tau_0}^{-\alpha} \quad (15)$$

$$\text{exponential} \quad srf(t) = \frac{1}{\tau_0} e^{-t/\tau_0} h(t) \quad (16)$$

where the linear response has the shortest effect and hyperbolic response has the longest effect on time series. We employ decaying response functions for two reasons. First, topics often become obsolete and cease being published in a short time period. Second, the shapes of response functions often bear additional information regarding impact and expectation of topic.

4.2 Topic deconvolution

Deconvolution is the opposite process of convolution (Gaikovitch 2004), which aims to recreate the original topical importance sequence. The Convolution theorem states that

the Fourier transformation of a time-domain convolution of two series is equal to the multiplication of their Fourier transformation in the frequency domain:

$$\mathbb{F}\{n(t)\} = \mathbb{F}\{e(t, r) * srf(t)\} = \mathbb{F}\{e(t, r)\} \cdot \mathbb{F}\{srf(t)\} \quad (17)$$

The problem now lies in how to integrate the temporal dynamics described above into our heterogeneous topic model and then introduce the fitting process to estimate its parameters. We encode the social response and topic importance by associating the Dirichlet parameters for each topic with a time-dependent function, which controls the popularity of the associated topics and the level of social response. Specifically, we let each dimension β_w in Dirichlet parameter β be associated with the following time-dependent function.

$$\beta_k(t) = f_k(t) = \mathbb{F}\{e(t, r)\} \cdot \mathbb{F}\{srf(t)\} \quad (18)$$

where $f_k(t)$ is the deconvolution model described in Section 4.2. However, if we naively associate β_k with f_k , the model may consider the starting point of time t for all topics is timestamp 0. In fact, different topics have different starting point t_0 . Thus we modify it into the following form:

$$f_k(t) = N_k + \mu(t - t_0^k)|t - t_0^k|^{q_k} * \mathbb{F}\{e(|t - t_0^k|, r)\} \cdot \mathbb{F}\{srf(|t - t_0^k|)\} \quad (19)$$

where t_0 is the starting timestamp of the topic, q_k indicates how quickly the topic would rise to the peak, and N_k is the noise level of the topic. The absolute value function guarantees that the time-dependent part is only active when t_0 is larger than t_0 . $\mu(t - t_0^k)$ is a boolean function that is 1 for $t > t_0$ and 0 otherwise. The intuition behind this equation is that the prior knowledge of each topic is fixed over time (by the “noise” level N_k). The crux of the problem is to estimate the values of these 3 hyper-parameters from the data.

Algorithm 2 Model fitting for topic deconvolution

Input : Input data, which includes: $w_{di}^{(k)}$, the number of words in Twitter feed di , $w_d^{(n)}j$, the number of Words in News feed dj , the number of topics T , and the Newton step parameter ξ

Output: $\phi = P(w_j|z_k)$, $\theta = P(z_k|d_i)$, and $\mathcal{A} = P(z_k|A_i)$

```

1 Random initialize the Gibbs Sampler
2  $t \leftarrow 0$ ; while  $t < MaxIteration$  do
3   E-step: For all feeds in all text streams, update topic assignments using (3) to (7)
4   M-step:
5     Update  $\alpha$  and  $\beta$  values through the method introduced in Hong et al. (2011)
6   for Each local and common topic do
7     Fit “temporal beta” function by using the parameters from the (19)
8     Re-estimate  $\beta$  values for topic  $t$  by using fitted function (19)
9   end
10 end
```

The procedures for parameter estimation is summarised in Algorithm 2. Generally, we integrate the deconvolution function with Gibbs Sampling into EM framework (e.g., similar to (Doyle and Elkan 2009)). In the E-step, we gather topic assignments and useful counts by Gibbs sampling using (3). In the M-step we optimise the proposed deconvolution functions to obtain the updated hyper-parameters for the next iteration. More specifically, the first step is to calculate the Dirichlet parameters β from word frequency observed from Gibbs Sampler. This can be done in several ways (Minka 2000). We use Newton’s method in this

step, where The step parameter ξ can be interpreted as a controlling factor of smoothing the topic distribution among the common topics. The second step is to use these β values to fit the deconvolution function (19) and then, use the parameters from the fitted deconvolution function as initial values to fit our temporal dynamic function (19).

5 Experiment and results

In particular, this paper aims to investigate three main research questions:

1. How to evaluate the performance of topic modelling techniques in the intertwined heterogeneous context?
2. How effective is our proposed heterogeneous topic model compared to other topic models in the above context?
3. Can temporal dynamics of each sources be exploited effectively to further enhance the performance of topic modelling?

5.1 Datasets and metrics

To construct two parallel datasets, we crawled 30 million tweet from 786,823 users from the first hour of April 1, 2015 to 720, the last hour of April 30, 2015 in a matter of one month in the region of Scotland. All these tweets follow one of the following rules, what we call, Scotland-related information centres. It consists of person names, places, and organisations that share information related to Scotland.

In addition, over the entire month, we crawled 224,272 unique resolved URLs. The News dataset is obtained through the Boilerpipe program.¹ The total dataset size is 4GB and essentially includes complete online coverage: we have all mainstream media sites plus 100,000 million blogs, forums, and other media sites. All our experiments are based on these two datasets.

For each web page, we collected (1) The title of the web page. (2) Text content of the web page (after removing tags, js, etc.) (3) Tweets linking to the page. For both the Twitter and the news media datasets, we first removed all the stop words. Next, the words with a document frequency of less than 10 and words that appeared in more than 70% of the tweets (news feeds) were also removed. Finally, for Twitter data, we further removed tweets with fewer than three words and all the users with fewer than 8 tweets.

The first metric we used is perplexity, which aims to measure how well a probability model can predict a new coming feed (Rosen-Zvi et al. 2004). Better generalization performance is indicated by a lower value over a held-out feed collection. The basic equation is:

$$perplexity = \exp\left(-\frac{\sum_{d=1}^D \sum_{i=1}^{N_d} \log p(w_{d,i} | \mathcal{D}^{\text{train}}, \alpha^{(n)}, \alpha^{(k)}, \beta)}{\sum_{d=1}^D N_d}\right) \quad (20)$$

where $w_{d,i}$ represents the i^{th} word in feed d . Note that the perplexity is defined by summing over the feeds.

¹<https://code.google.com/p/boilerpipe/>

However, the output of different topic models may be significantly different. For example, the outputs of Source-LDA are document-topic distribution and topic-term distribution, while HTM generates topic-term and author-topic distributions:

$$P_{\text{LDA}}(w_{d,i}|\mathcal{D}^{\text{train}}, \alpha^{(n)}, \alpha^{(k)}, \beta) = \sum_{t=1}^T \theta_{dt} \eta_{tw} \quad (21)$$

$$P_{\text{ATM}}(w_{d,i}|\mathcal{D}^{\text{train}}, \alpha^{(n)}, \alpha^{(k)}, \beta) = \sum_{t=1}^T \phi_{ta} \eta_{tw} \quad (22)$$

As a result, the outputs of different models are incomparable. Recalling our first research questions at the beginning of Section 5, to make a fair comparison, we customise the perplexity metric into the heterogeneous context

$$\text{perplexity}(T) = \exp\left(-\frac{\sum_{t=1}^T \log p(w_{d,i}|\mathcal{D}^{\text{train}}, \alpha^{(n)}, \alpha^{(k)}, \beta)}{\sum_{d=1}^D N_d}\right) \quad (23)$$

where $p(w_{d,i}|\mathcal{D}^{\text{train}}, \alpha^{(n)}, \alpha^{(k)}, \beta)$ equals to η_{tw} , and T is the total number of topics. The proposed topic perplexity can explain how well each topic model can predict an unseen topic. A low value of perplexity is an indicator of good performance for the topic model that is being evaluated.

Another common metric to evaluate the performance of topic model is entropy (Li et al. 2004), which denotes the expected value of information contained in the message. Similar to the perplexity, we define the following equation to compute the topic entropy.

$$\text{entropy}(T) = \exp\left(-\frac{\sum_{d=1}^D \frac{1}{N_d} \sum_{i=1}^{N_d} \log p(z|\mathcal{D}^{\text{train}}, \alpha^{(n)}, \alpha^{(k)}, \beta)}{N}\right) \quad (24)$$

Again, the smaller value of the entropy measure, the better are the topics since it indicates a better discriminative power.

This paper compares four topic models:

1. Source-LDA, which identifies the latent topics by leveraging the distribution of each of the individual sources. Notice that we use Twitter as the main source as it has reported in Ghosh and Asur (2013) that choosing Twitter as the main source can achieve an superior performance over other options.
2. Heterogeneous Topic Model (HTM), the model we described in Section 3 which applies Author Topic Model (ATM) on Twitter dataset and LDA on News dataset, as a result the properties of each source can be preserved.
3. Deconvolution Model (DM), which simply integrate the deconvolution model described in Section 4.2 into Author Topic Model (ATM) of Twitter dataset.
4. Heterogeneous Topic Model with Temporal Dynamics (HTMT) (cf. Section 4.2) which incorporate the deconvolution model into HTM.

5.2 Parameter setting

We randomly sample 80% of the data as the training data and use the remaining 20% as the test data. All models are trained on the same training set and evaluated using the same test set. In the training phase we obtain topic-term distribution, the number of topics, and all other hyper-parameters. In the testing phase we fix them and make 200 Gibbs-sampling

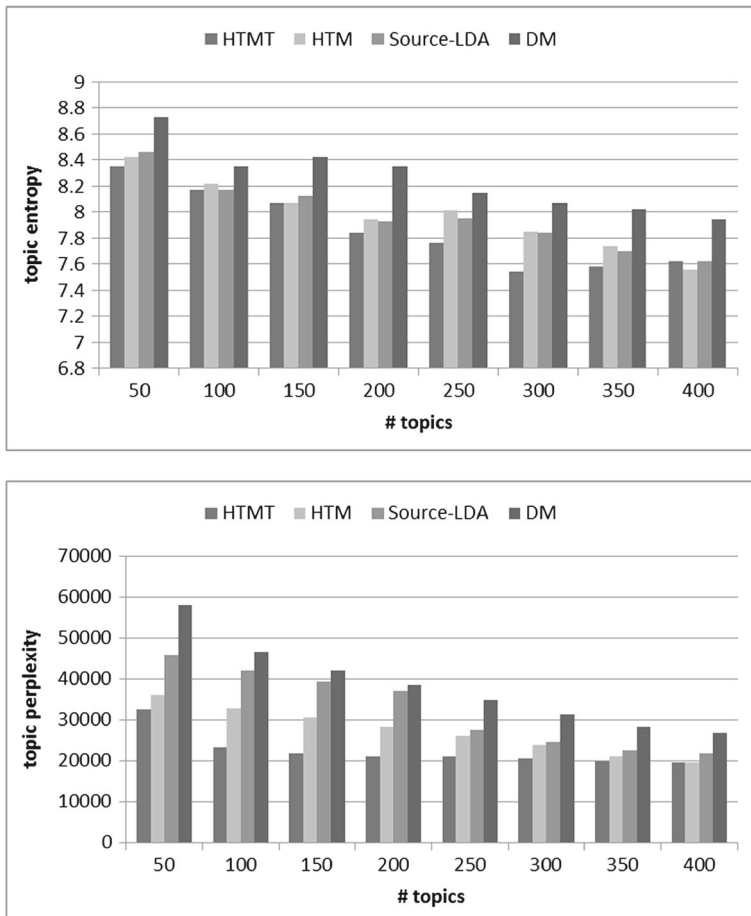


Fig. 2 This figure shows the topic entropy and perplexity of topic models with varying number of topics T with a linear decay function

iterations for each feed in the test set, obtaining θ_{dt} and ϕ_{ta} . It is well known that in general we need to use different topics for different datasets to achieve the best topic modelling effect. Hence, we tried the topic models with different values of T . The source code is made open to the research community as online supplementary material.²

For DM, we only use twitter messages for clustering with no additional news information. For HTM, we use symmetric Dirichlet priors in the LDA estimation with $\alpha = 50/T$ and $L = 0.01$, which are common settings in the literature. For HTMT, both HTM and Deconvolution Model need to be tuned. Usually, deconvolution with small decay parameters is the best setting. However, only a high level of deconvolution helps to restore the original topic importance, as well as to depict its temporal dynamics using the desired model. Higher

²<https://github.com/whitezhang/TopicModel>

than necessary deconvolution may lead to smaller topics (which are caused by the prior larger neighbours), so we need to apply the lowest deconvolution possible, but in the same time maintain the adequate level of deconvolution using the parameter estimation method as described in the next section. The parameter settings of Deconvolution Model were set to be identical to those in Tsytarau et al. (2014).

Another important parameter for topic modelling is the number of topic T . To find the optimal value for T , we experiment different topic models on the training dataset. In Hong et al. (2011), the best performance is achieved when $K = 50$. As shown in Figs. 2, 3 and 4, however, it is clear that the optimal performance achieved when the number of topic is equal to 300 for both metrics. One possible explanation is that our dataset is substantially larger than their dataset, which requires larger amount of topics to model it.

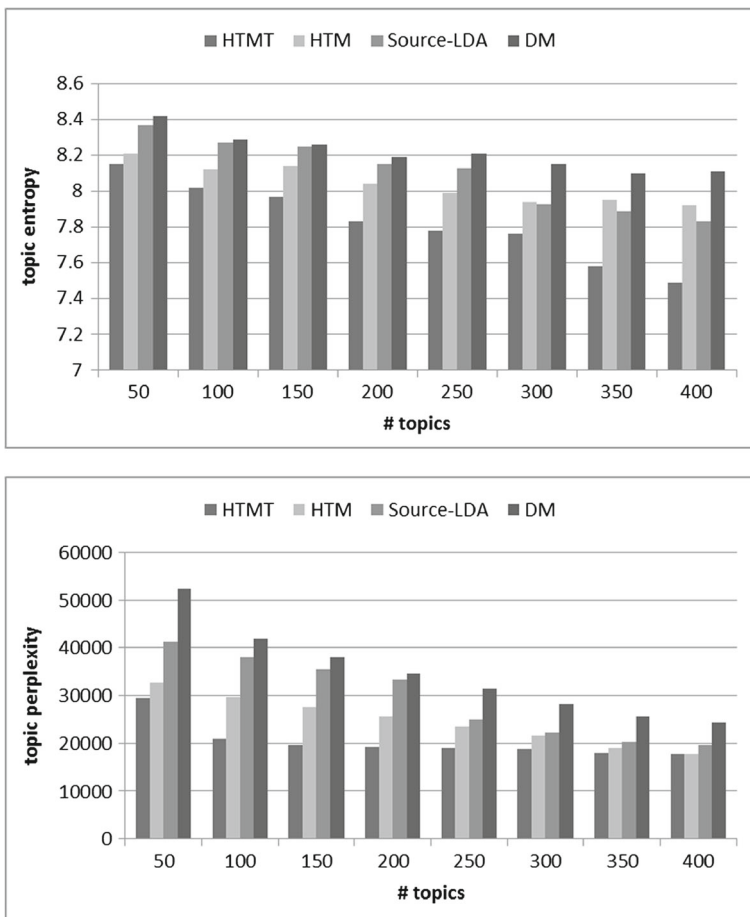


Fig. 3 This figure shows the topic entropy and perplexity of topic models with varying number of topics T with a hyperbolic decay function

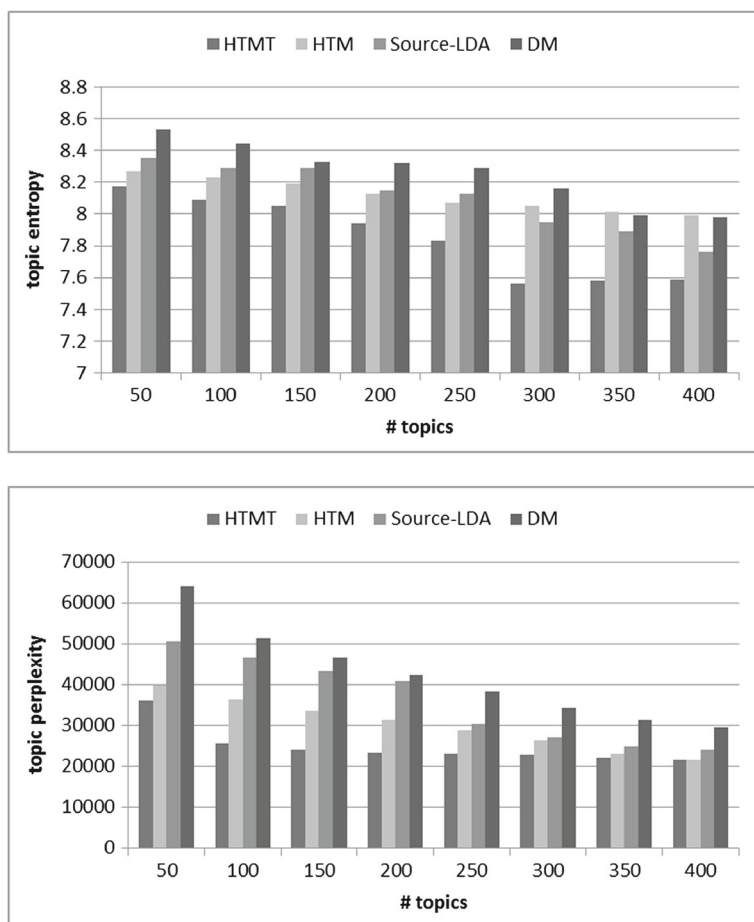


Fig. 4 This figure shows the topic entropy and perplexity of topic models with varying number of topics T with a exponential decay function

From the training stage, one can also see that Heterogeneous Topic Model (HTM) brings substantial performance gain compared to the state-of-the-art approach, the Source-LDA. However, the proposed Deconvolution Model (DM) can only achieve a comparable performance to Source-LDA when applying alone. We conjecture that this is because DM is largely dominated by the local topics of Twitters (due to its high volume). More importantly, when combining DM with HTM, we see that our proposed approach HTMT, which considers both the heterogeneous properties of each source and the interwind temporal dynamics, outperforms all the other approaches, irrespective of the topic number and the evaluation metrics.

Table 2 The performance of different methods on (a) Twitter and (b) News datasets (-*-* and -* indicate the statistical significance of performance decrease from that of SGTM with p-value < 0.01 and p-value < 0.05, respectively)

	Source-LDA	HTM	DM	HTMT
(a) Twitter				
<i>entro</i>	11.375-*-*	10.596-*	9.642-*	8.798
<i>perp</i>	7269.529-*-*	5924.371-*-*	4943.582-*	4136.335
(b) News				
<i>entro</i>	9.374-*-*	8.647-*-*	8.551-*	8.295
<i>perp</i>	6654.247-*-*	5753.293-*-*	4439.986-*	3782.961

5.3 Topic model analysis and case study

A comparison using the paired t-test is conducted for HTM, DM, and Source-LDA over HTMT, as shown in Table 2. It is clear that HTMT outperform all baseline methods significantly on both Twitter and News dataset. This indicates that combining heterogeneous source with temporal dynamics significantly improved the performance of topic modelling.

In order to visualise the hidden topics and compare different approaches, we extract topics from training datasets using Source-LDA, HTM, DM, and HTMT. Since both sources consist of a mixture of Scottish subjects, it is interesting to see whether the extracted topics could reflect this mixture. We select a set of topics with the highest value of $p(z|d)$, which represents the average of the expected value of topic k appearing in a feed on both collections. For each topic in the set, we then select the terms of the highest $p(w|z)$ value. Notice that all these models are trained in an unsupervised fashion with $k = 300$, all the other settings are the same with the above experiments, the decay function is set as hyperbolic since it exhibits the best performance on the test dataset (cf. Figs. 2, 3 and 4). The text have been preprocessed by case-folding and stopwords-removal (using a standard list of 418 common words). Shown in Table 3 are the most representative words of topics generated by Source-LDA, HTM, DM, and HTMT respectively. For topic 1, although different models select slightly different terms, all these terms can describe the corresponding topic to some extent. For topic 2 (Glasgow), however, the words “commercial” and “jobs” of HTMT are more telling than “food” derived by Source-LDA, and “beautiful” and “smart” derived by HTM. Similar subtle differences can be found for the topic 3 and 4 as well. Intuitively, HTMT selects more related terms for each topic than other methods, which shows the better performance of HTMT by considering both the heterogeneous structure and the temporal dynamics information. This observation answered the last research question, where our HTMT framework supersedes HTM by providing more valuable and reinforced information.

5.4 Analysis on temporal dynamics

Hashtag, a type of community convention that starts with a “#” sign, have been extensively used as annotations to represent events and topics on Twitter. We select several hashtag that can act as indicators for certain events where each hashtag is clearly associated to some events in April, 2015. More specifically, we choose #Scotrail for “Scottish railway”, and wish to see whether the events can be discovered by different models and how well these models can be presented. We believe these hashtag represent a large range of social events and therefore are representative. A natural question is whether the model can identify topics that reflect the events behind the hashtags. We map hashtags onto the topics obtained by the

Table 3 The representative terms generated by Source-LDA, HTM, DM, and HTMT models

	Topic 1 (VoteSNP)		Topic 2 (Glasgow)		Topic 3 (Scotrail)		Topic 4 (Loch Ness)	
HTM Source-LDA	VoteSNP	wish	Glasgow	Scotish	Scotrail	transport	loch	<u>catfish</u>
	Scotish	support	city	<u>food</u>	Scotish	franchise	nessie	<u>marathon</u>
	voting	UK	UK	Scotland	route	<u>company</u>	monster	deep
	party	<u>first</u>	Scotland	people	train	disruption	inverness	<u>food</u>
	Scotland	member	council	project	rail	lines	visit	life
	VoteSNP	<u>leaflet</u>	Glasgow	merchant	Scotrail	transport	loch	water
	Scotland	support	Edinburgh	investment	Scotish	rail	lochness	beautiful
	Scotish	UK	Scotland	Scotland	station	fares	monster	tour
	party	delivering	Scotish	<u>beautiful</u>	route	<u>journey</u>	inverness	<u>picture</u>
	voting	<u>campaign</u>	council	<u>smart</u>	train	Edinburgh	<u>celebrity</u>	<u>lodge</u>
DM	VoteSNP	wish	Glasgow	Scotish	Scotrail	transport	loch	<u>lodge</u>
	Scotland	<u>love</u>	city	London	Scotish	franchise	lochness	tour
	Scotish	support	UK	council	route	company	nessie	highland
	voting	everyone	Scotland	project	train	Edinburgh	monster	<u>best</u>
	UK	party	Edinburgh	merchant	<u>ticket</u>	fare	inverness	<u>life</u>
HTMT	VoteSNP	<u>labour</u>	Glasgow	Scotish	Scotrail	rail	loch	<u>hideaway</u>
	Scotland	support	Scotland	<u>commercial</u>	Scotish	<u>delay</u>	nessie	highland
	Scotish	everyone	Edinburgh	Scotland	transportation	<u>signal</u>	monster	<u>deep</u>
	Glasgow	Edinburgh	city	<u>job</u>	train	disruption	inverness	<u>uncover</u>
	party	<u>action</u>	council	investment	<u>smoker</u>	franchise	visit	<u>home</u>

The terms are vertically ranked according to the probability $P(w|z)$
Bold and underlined data indicates the unique words that are captured by our model

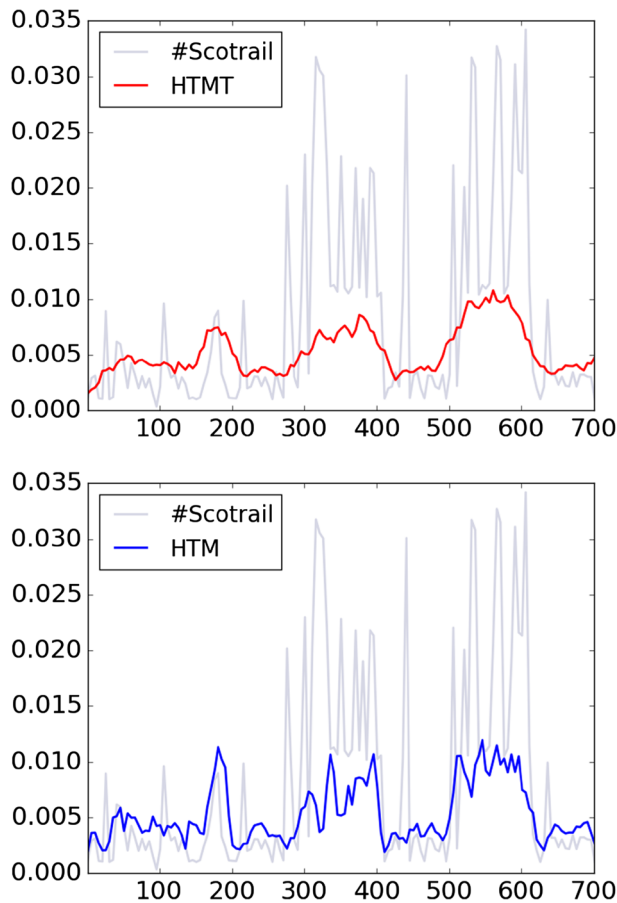


Fig. 5 Comparison of the probability $p(t|z)$ between the HTMT (top) and the HTM (bottom) models against the hashtag Scotrail. X-axis is the hour number, Y-axis is the probability

models and top ranked terms in these topics are examined to see whether the corresponding terms have any relationships with the underlying events.

To map the hashtags, we calculate the following probability $p(z|w) = \frac{p(w|z)p(z)}{\sum_{z'} p(w|z')p(z')}$ (Hong et al. 2011) where $p(w|z)$ is provided by the trained models and $p(z)$ can be easily estimated by the counts. Intuitively, this probability tells us how likely a topic is to be selected, given the term. We can then compare the time-series of topics and hashtags to determine whether they are similar. Our hypothesis is that if they look similar on the time series, the topic may be good choices for explaining the events behind the hashtags. Notice that we are not seeking the exact match here since the topics have many more terms than a single hashtag and it may explain multiple events. Moreover, we transform the volumes into probabilities. We plot the time series of hashtags and the time series of the selected topics in Figs. 5, 6 and 7.

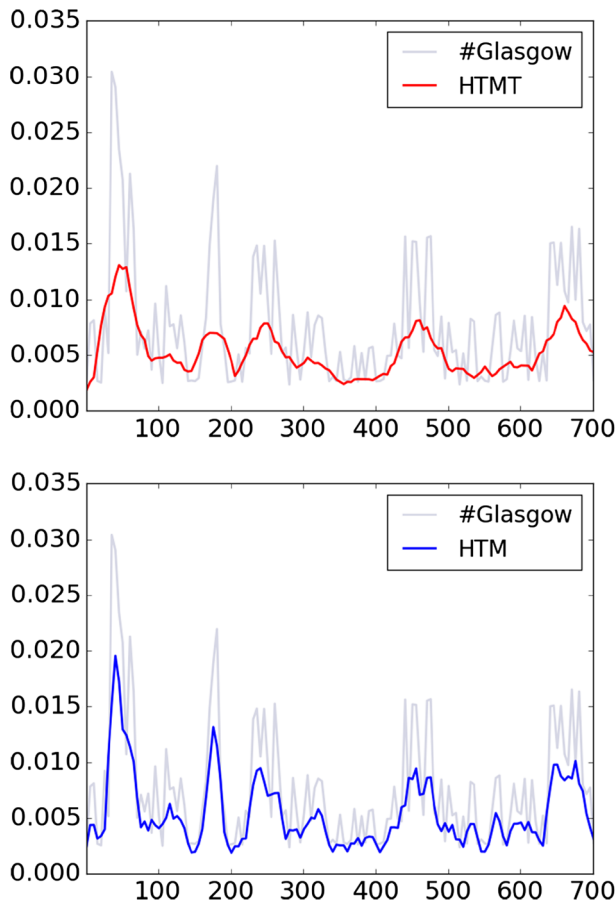


Fig. 6 Comparison of the probability $p(t|z)$ between the HTMT (top) and the HTM (bottom) models against the hashtag Glasgow. X-axis is the hour umber, Y-axis is the probability

From the result, HTMT and HTM (red and blue curves) both smooth out the local fluctuations of the topics (gray curves) from the hashtags shown, while preserving the sharp peaks that may indicate a significant change of content in Twitter. Moreover, HTM tend to “over-fit” the multiple spikes in the occurrence of “Scotrail” between 300 and 400 of hours. Also, HTMT can better match the peaks of hashtags, indicating that the method can better reflect real events. This may owe to the fact that the hyper-parameters β in HTMT are governed by the time-dependent functions of both sources, where the rise and fall of these values may give good hints for the model to assign topic to words, leading to a improved performance on temporal dynamics, and provides additional answers to the last research question at the beginning of this section.

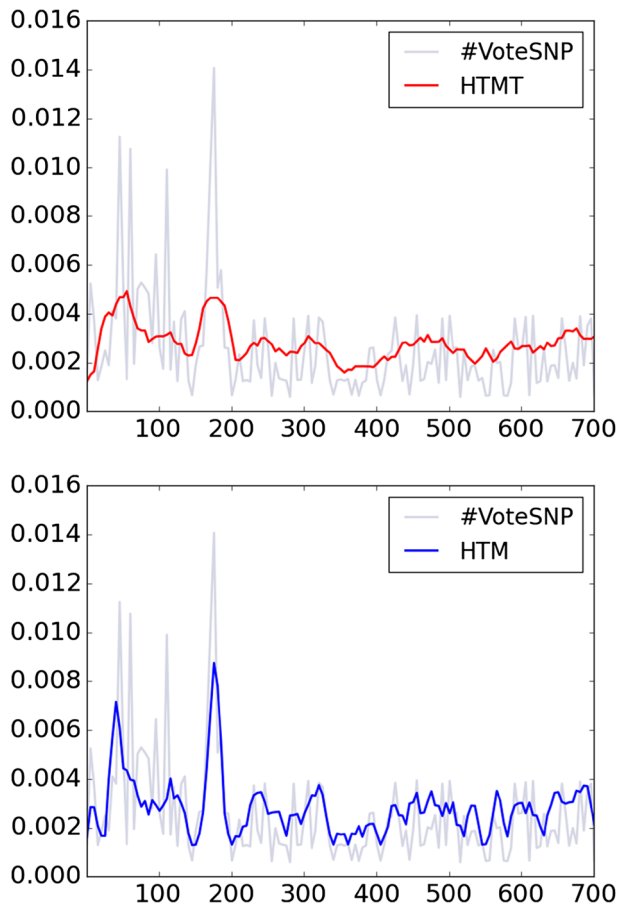


Fig. 7 Comparison of the probability $p(t|z)$ between the HTMT (top) and the HTM (bottom) models against the hashtag VoteSNP. X-axis is the hour number, Y-axis is the probability

6 Conclusion

Mining topics from heterogeneous sources is still a challenge, especially when the temporal dynamics of the sources are intertwined. In this paper, we aim to automatically analyse multiple correlated sources with their corresponding temporal behaviour. The new model goes beyond the existing Source-LDA models because (i) it blends several topic models while preserving the characteristic of each source; (ii) it associates each topic with a deconvolution function that characterise its topic importance and social response over time, which enables our topic model to capture some of the hidden contextual information of feeds.

There are some interesting future work to be continued. First, it will be interesting to investigate more complex aspects related to the temporal dynamics of data stream, e.g., the sentiment shift. Additionally, in order to model and gain insight from real events, topics can be linked with a group of named entities such that each topic can be largely explained by these entities and their relations in the knowledge graphs, e.g., Freebase.³

³<https://www.freebase.com/>

Acknowledgements We thank the anonymous reviewer for their helpful comments. We acknowledge support from the EPSRC funded project named A Situation Aware Information Infrastructure Project (EP/L026015) and from the Economic and Social Research Council [grant number ES/L011921/1]. This work was also partly supported by NSF grant #61572223. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the view of the sponsor.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

In this section, we show how to derive the Gibbs Sampler (2) for HTM (cf. Section 3.1). By integrating Φ , \mathcal{B} and x , we get:

$$p(z, w^{(k)} | \mathcal{A}, \alpha^{(k)}, \eta) = \int_x \int_{\Phi} \int_{\mathcal{B}} p(z|x, \Phi) p(w^{(k)}|z, \mathcal{B}) p(x|\mathcal{A}) p(\Phi|\alpha^{(k)}) p(\mathcal{B}|\eta) dx d\Phi d\mathcal{B} \quad (25)$$

$$= \int_x \int_{\Phi} \int_{\mathcal{B}} \left\{ \left[\prod_{i=1}^N p(z_{di} | \phi_{x_{di}}) \right] \left[\prod_{d=1}^D \prod_{i=1}^{N_d} p(x_{di} | a_d) \right] p(\Phi | \alpha^{(k)}) p(\mathcal{B} | \eta) \right\} dx d\Phi d\mathcal{B} \quad (26)$$

$$= \int_x \int_{\Phi} \int_{\mathcal{B}} \left\{ \left[\prod_{a=1}^A \prod_{t=1}^T \beta_{ta}^{C_{ta}^{TA}} \right] \left[\prod_{d=1}^D \left(\frac{1}{A_d} \right)^{N_d} \right] \left[\prod_{t=1}^T \left(\frac{\Gamma(W\eta)}{\Gamma(\eta)^W} \prod_{w=1}^W \phi_{wt}^{\eta_w-1} \right) \right] \right. \\ \left. \left[\prod_{a=1}^A \left(\frac{\Gamma(T\alpha)}{\Gamma(\alpha)^T} \prod_{t=1}^T \beta_{ta}^{\alpha_t^{(k)}-1} \right) \right] \right\} dx d\Phi d\mathcal{B} \quad (27)$$

$$= \int_x \int_{\Phi} \int_{\mathcal{B}} \left\{ \left[\prod_{a=1}^A \prod_{t=1}^T \beta_{ta}^{C_{ta}^{TA}} \right] \left[\prod_{d=1}^D \left(\frac{1}{A_d} \right)^{N_d} \right] \left[\prod_{t=1}^T \prod_{w=1}^W \beta_{wt}^{C_{wt}^{WT} + \eta_w - 1} \right] \right. \\ \left. \left[\prod_{a=1}^A \prod_{t=1}^T \phi_{ta}^{C_{ta}^{TA} + \alpha_t^{(k)} - 1} \right] \right\} dx d\Phi d\mathcal{B}, \quad (28)$$

$$= \left[\prod_{a=1}^A \prod_{t=1}^T \beta_{ta}^{C_{ta}^{TA}} \right] \int_{x: \frac{1}{A_d}} \left[\prod_{d=1}^D \left(\frac{1}{A_d} \right)^{N_d} \right] dx \cdot \frac{1}{A_d} \int_{\mathcal{B}} \left[\prod_{t=1}^T \prod_{w=1}^W \beta_{wt}^{C_{wt}^{WT} + \eta_w - 1} \right] d\mathcal{B} \int_{\Phi} \\ \left[\prod_{a=1}^A \prod_{t=1}^T \phi_{ta}^{C_{ta}^{TA} + \alpha_t^{(k)} - 1} \right] d\Phi \quad (29)$$

$$= \left[\prod_{a=1}^A \prod_{t=1}^T \beta_{ta}^{C_{ta}^{TA}} \right] \left[\prod_{d=1}^D \left(\frac{1}{A_d} \right)^{N_d+1} \right] \left[\prod_{a=1}^A \frac{\prod_{t=1}^T \Gamma(C_{ta}^{TA} + \alpha_t^{(k)})}{\Gamma(\sum_{t'} C_{t'a}^{TA} + T\alpha^{(k)})} \right] \\ \left[\prod_{t=1}^T \frac{\prod_{w=1}^W \Gamma(C_{wt}^{WT} + \eta_w)}{\Gamma(\sum_{w'} C_{w't}^{WT} + W\eta)} \right] \quad (30)$$

where C_{wt}^{WT} denotes the number of times that word w in the corpus is assigned to t^{th} topic and C_{ta}^{TA} is the number of times that topic t is assigned to author a . Then we can use the same approach as (25) to have:

$$p(z_{-di}, w_{-di}^{(k)} | \mathcal{A}, \alpha^{(k)}, \eta) = \left[\prod_{a=1}^A \prod_{t=1}^T \beta_{ta}^{C_{ta}^{TA}} \right] \left[\prod_{d=1}^D \left(\frac{1}{A_d} \right)^{N_d+1} \right] \left[\prod_{a=1}^A \frac{\prod_{t=1}^T \Gamma(C_{ta}^{TA} + \alpha_t^{(k)})}{\Gamma(\sum_{t'} C_{t'a}^{TA} + T\alpha^{(k)})} \right] \left[\prod_{t=1}^T \frac{\prod_{w=1}^W \Gamma(C_{wt}^{WT} + \eta_w)}{\Gamma(\sum_{w'} C_{w't}^{WT} + W\eta)} \right] \quad (31)$$

Using (25) and (31), the following equation can be derived:

$$\frac{p(z, w^{(k)} | \mathcal{A}, \alpha^{(k)}, \eta)}{p(z_{-di}, w_{-di}^{(k)} | \mathcal{A}, \alpha^{(k)}, \eta)} = \frac{C_{wt, -di}^{WT} + \eta_w}{\sum_{w'} C_{w't, -di}^{WT} + W\eta} \cdot \frac{C_{ta, -di}^{TA} + \alpha_t^{(k)}}{\sum_{t'} C_{t'a, -di}^{TA} + T\alpha^{(k)}} \quad (32)$$

Similarly, $p(z, w^{(n)} | \alpha^{(n)}, \eta)$ can be calculated as follows:

$$p(z, w^{(n)} | \alpha^{(n)}, \eta) = p(w^{(n)} | z, \eta) p(z | \alpha^{(n)}) \quad (33)$$

where $p(w^{(n)} | \alpha^{(n)}, \eta)$ and $p(z | \alpha^{(n)})$ can be obtained according to Heinrich (2005) as follows:

$$p(w^{(n)} | z, \eta) = \prod_{t=1}^T \frac{\Delta(n_t + \eta)}{\Delta(\eta)}, n_t = \{n_t^{(w)}\}_{t=1}^V. \quad (34)$$

$$p(w^{(n)} | z, \alpha^{(n)}) = \prod_{d=1}^D \frac{\Delta(n_d + \alpha^{(n)})}{\Delta(\alpha^{(n)})}, n_d = \{n_d^{(t)}\}_{t=1}^T \text{endaligned} \quad (35)$$

References

- Alsumait, L., Barbará, D., & Domeniconi, C. (2008). Online lda: Adaptive topic model for mining text streams with application on topic detection and. *ICDM'08*.
- Asur, S., Huberman, B.A.s., Szabo, G., & Wang, C. (2011). Trends in social media: Persistence and decay. Available at SSRN 1755748.
- Blei, D.M., & Lafferty, J.D. (2006). Dynamic topic models. In *ICML '06* (pp. 113–120).
- Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, P.K. (2010). Measuring user influence in twitter: the million follower fallacy. *ICWSM '10*, 10(10–17), 30.
- Crane, R., & Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41), 15649–15653.
- Doyle, G., & Elkan, C. (2009). Accounting for burstiness in topic models. *ICML '09*.
- Dubey, A., Hefny, A., Williamson, S., & Xing, E.P. (2013). A nonparametric mixture model for topic modeling over time.
- Fiscus, J.G., & Doddington, G.R. (2002). Topic detection and tracking. chapter Topic Detection and Tracking Evaluation Overview, pp. 17–31.
- Gaikovich, K.P. (2004). Inverse problems in physical diagnostics. Nova Publishers.
- Ghosh, R., & Asur, S. (2013). Mining information from heterogeneous sources: A topic modeling approach. In *Proc. of the MDS Workshop at the 19th ACM SIGKDD (MDS-SIGKDD '13)*.
- Heinrich, G. (2005). Parameter estimation for text analysis. Technical report.
- Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, 45, 256–269.
- Hong, L., Dom, B., Gurumurthy, S., & Tsioutsoulis, K. (2011). A time-dependent topic model for multiple text streams.

- Hong, L., Yin, D., Guo, J., & Davison, B.D. (2011). Tracking trends: incorporating term volume into temporal topic models.
- Kirkeby, O., Nelson, P., Hamada, H., Orduna-Bustamante, F., et al. (1998). Fast deconvolution of multichannel systems using regularization. *IEEE Transactions on Speech and Audio Processing*, 6(2), 189–194.
- Lau, J.H., Collier, N.s., & Baldwin, T. (2012). On-line trend analysis with topic models: \# twitter trends detection topic model online. In *COLING* (pp. 1519–1534).
- Leskovec, J., Backstrom, L., & Kleinberg, J. (2009). Meme-tracking and the dynamics of the news cycle.
- Li, T., Ma, S., & Ogihara, M. (2004). Entropy-based criterion in categorical clustering.
- Mallat, S. (1999). A wavelet tour of signal processing. Academic press.
- Masada, T., Fukagawa, D., Takasu, A., Hamada, T., Shibata, Y., & Oguri, K. (2009). Dynamic hyperparameter optimization for bayesian topical trend analysis. *CIKM '09*.
- Miller, J.W., & Alleva, F. (1996). Evaluation of a language model using a clustered model backoff. volume 1 of *ICSLP' 96*, pages 390–393.
- Minka, T. (2000). Estimating a dirichlet distribution.
- Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2004). The author-topic model for authors and documents.
- Tsytsarau, M., Palpanas, T., & Castellanos, M. (2014). Dynamics of news events and social media reaction.
- Wang, C., Blei, D., & Heckerman, D. (2012). Continuous time dynamic topic models. arXiv:[1206.3298](https://arxiv.org/abs/1206.3298).
- Wang, X., & McCallum, A. (2006). Topics over time: a non-markov continuous-time model of topical trends.
- Zhai, C.X., Velivelli, A., & Yu, B. (2004). A cross-collection mixture model for comparative text mining. In *KDD '04* (pp. 743–748).
- Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., & Li, X. (2011). Comparing twitter and traditional media using topic models. In *ECIR'11* (pp. 338–349).